DOCUMENT RESUME

ED 068 487 TM 001 823

TITLE Aptitude Testing: A Critical Examination of the Differential Aptitude Tests, Alternative Batteries, and Problems in Prediction.

INSTITUTION Toronto Board of Education (Ontario). Research Dept.

NOTE 20p.; Report no. 17

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Academic Aptitude; Achievement; *Aptitude Tests; Manuals; Motivation; Performance; Prediction; *Predictive Ability (Testing); Student Evaluation; Student Interests; Test Reliability; *Test Reviews; Test Validity; *Vocational Aptitude; Vocational Interests

ABSTRACT

In addition to a review of the Differential Aptitude Tests (DAT), a number of other aptitude tests are examined. They are: (1) Flanagan Aptitude Classification Tests, (2) Holzinger-Crowder Uni-Factor Tests, (3) Employee Aptitude Survey, (4) Revised Minnesota Paper Form Board Test, (5) Minnesota Clerical Test, and (6) Turse Clerical Aptitudes Test. The results suggest that aptitude tests have been useful instruments in predicting general scholastic aptitudes but have not been as successful in predicting technical aptitudes. Possible reasons for this finding may be that (1) a student's high interest does not necessarily indicate a high aptitude, and (2) the inability to measure every aspect of the achievement process. It is concluded that further research is necessary for aptitudes to be identified and measured more accurately. It is felt that much of a test's usefulness is the adequacy of its manual. With the exception of the DAT manual, most seemed to lack presenting adequate norms, their source, and appropriate validity and reliability data. It is suggested that (1) present multifactor batteries do not adequately differentiate aptitudes, 2) adequate guidance will depend on supplementary information, and (3) student motivation requires more attention and measurement. (JS)

ED 068487

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

#17 Aptitude Testing: A Critical Examination of the Differential Aptitude Tests, Alternative Batteries, and Problems in Prediction
Out-of-Print

FOR

APTITUDE TESTING: A CRITICAL EXAMINATION OF THE DIFFERENTIAL APTITUDE TESTS, ALTERNATIVE BATTERIES, AND PROBLEMS IN PREDICTION

TM 001 823

RESEARCH SERVICE

issued by the
Research Department

FILMED FROM BEST AVAILABLE COPY

THE BOARD OF EDUCATION FOR THE CITY OF TORONTO

TABLE OF CONTENTS

		Page No.
I	INTRODUCTION.....	1
II	THE DIFFERENTIAL APTITUDE TESTS.....	2
III	OTHER MULTIFACTOR APTITUDE TESTS.....	6
	Flanagan Aptitude Classification Tests Holzinger-Crowder Uni-Factor Tests Employee Aptitude Survey	
IV	SPECIALIZED TESTS.....	10
	The Revised Minnesota Paper Form Board Test The Minnesota Clerical Test Purse Clerical Aptitudes Test	
V	APTITUDES VERSUS INTERESTS.....	14
VI	THE STUDENT AS A SOURCE OF ERROR IN PREDICTING ACHIEVEMENT.....	15
VII	GENERAL CONCLUSION.....	16
	REFERENCES.....	17

APTITUDE TESTING: A CRITICAL EXAMINATION OF THE DIFFERENTIAL APTITUDE TESTS, ALTERNATIVE BATTERIES, AND PROBLEMS IN PREDICTION

I INTRODUCTION

The value of any psychological test is governed primarily by the care with which it has been constructed, its reliability and validity, and the manner in which it is used. Its validity depends on the criterion measures to be used. Its usefulness depends partly on the adequacy of the norms. The test, in effect, provides the tester with the equivalent of a highly structured interview that is designed to give maximum information in a given period of time.

II THE DIFFERENTIAL APTITUDE TESTS

The Differential Aptitude Tests (DAT), the first multifactor battery available (1947), is one of the best and most widely accepted of such batteries. The DAT manual provides comparatively adequate information on the various aspects of this battery's construction. The manual contains detailed data based on an extensive geographical sampling. Validity data are copiously reported, and detailed norms are provided by grade and sex with score profiles at various educational levels.

The validity coefficients given in the manual are statistical expressions of correlations between particular tests and criterion measures such as course grades, other achievement tests, and intelligence tests. One cannot ask for a general statement of test validity since a variety of criterion measures are involved in a multifactor test. The manual emphasizes that validity is "specific", each coefficient is relevant to a given sample and is based on a given criterion measure. While the norms are based on an impressive sample -- 47,000 -- many of the validity coefficients are based on relatively small samples. Interpretation of an individual's profile should, therefore, be based on the validity tables appropriate to the given grade level, sex and course, rather than on the generalities expressed in the manual. In view of the many factors that must be taken into consideration, the process of interpretation presents difficulties in deciding which validity coefficients are appropriate for the local situation.

Reliability data indicate that scores made on this battery will not significantly fluctuate over time. Score profiles between first,

second or third testing will remain constant. Reliability coefficients between Form A and Form B range from .70 to .94.

Bennett, Seashore, and Wesman (1956) summarize research on the long-term effectiveness of this test. Generally favourable results are reported in a wide variety of investigations, dealing with different levels and subject areas. This research summary is noted in the manual.

Studies conducted by Wolking (1955) and Vineyard (1958) have revealed that the DAT is best at predicting science and English grades. This statement is applicable to both sexes since validity differences between them were small. The Numerical Ability (NA) test proved to be the single most valid test of academic success, while the Verbal Reasoning (VR) test placed a close second. Together (VR + NA), these tests were better predictors of college success than the remaining tests of the DAT battery.

Correlations between VR + NA tests and several intelligence tests were also very high. Thus, the manual points out that the VR and NA tests measure "what is measured by intelligence and scholastic aptitude tests."

The DAT battery has been examined as a predictor in engineering success (Berdie, 1951). Consistently high significant correlations were found between high school rank, grades, and the Numerical Ability test. Language Usage, Mechanical Reasoning, and Abstract Reasoning tests proved to be of little predictive value.

After extensive examination of the manual, the follow-up studies, and the available literature, several conclusions can be drawn about the DAT:

(a) Comparatively speaking, the DAT manual stands far ahead of the majority of test manuals. The purpose of each test is well described, and adequate substantiation by statistical evidence is presented. Considerable effort has been devoted to making this manual informative and the procedures efficient. Further assistance is available in a casebook that outlines 30 sample score profiles and how they were interpreted. (Bennett, Seashore, and Wesman, 1951)

(b) Validity data presented in the manual are more adequate with respect to quantity than usefulness. A considerable time expenditure is required to wade through the overloaded and complex tables or charts. The tester, furthermore, has no way of knowing exactly what data are most appropriate for local interpretation.

(c) The DAT battery is more effective in predicting academic success, particularly in English and science courses. The Verbal Reasoning and Numerical Ability tests prove to be the most valuable predictors in this area.

(d) A certain degree of learning capacity is required in any course of study. While the DAT battery is valuable in that it provides a measure of scholastic potential, it does not differentiate aptitudes satisfactorily.

Individual tests demonstrating the highest correlation to specific course grades might be combined to provide "clusters" which relate to appropriate vocational fields. This procedure would reduce the number of tables, categorize the dominant abilities for given courses and vocations, and thereby facilitate profile interpretation. As yet, there have been no reports assessing the merits of this "cluster" principle.

(e) Greater use of expectancy tables, as described in the manual, may provide additional clarification in test interpretation. Such tables may be particularly useful in translating a profile to the student. Essentially, test scores are converted into "chances of success" in a given course or vocation. Research data are rarely published using this procedure.

(f) A local study with local norms and well-defined criterion measures would provide the most effective norms for the DAT battery. Such a study would indicate the level of confidence with which predictions could be made in the local system.

In summarizing then, several steps must be taken to achieve maximum predictive validity from the DAT. Local norms must be developed and appropriate criteria clearly identified for each group to be regularly tested. Such steps will increase the confidence with which predictions can be made about students' success in different programmes.

III OTHER MULTIFACTOR APTITUDE TESTS

This section is devoted to a brief examination of various aptitude tests which attempt to serve the same purpose as the DAT. Differences will be found primarily in age groups for which the test has been designed, in factor loadings, total testing time required, number and kind of subtests in each battery, extent of validity data and follow-up studies provided in the manual, and in the number of validated occupational categories furnished for purposes of interpretation.

The Flanagan Aptitude Classification Tests (FACT) 1951-56

Flanagan (1957) has taken the middle road between the "job elements" approach and the factor analysis approach. He analyzed the apparent abilities required of a great variety of occupations. These abilities were filtered by means of factor analysis, into 21 "critical job elements". Extensive studies with various subtests resulted in a final battery containing 21 tests. These tests are:

1. Inspection
2. Coding
3. Memory
4. Precision
5. Assembly
6. Scales
7. Co-ordination
8. Judgment and Comprehension
9. Arithmetic
10. Patterns
11. Components
12. Tables
13. Mechanics
14. Expression
15. Vocabulary
16. Reasoning
17. Planning
18. Ingenuity
19. Alertness
20. Tapping
21. Carving

} performance tests

While the FACT battery purports to measure different abilities, a quick perusal of the tests suggests distinctive factor loadings of abilities as shown below:

5 tests (1, 2, 4, 6, 9) involve speed and accuracy of a clerical nature

4 tests (5, 10, 11, 21) involve ability to deal with spatial relations

4 tests (8, 15, 16, 17) involve verbal reasoning ability

2 tests (14, 15) involve ability in language usage

4 tests (11, 17, 18, 19) involve the facility for abstract reasoning

3 tests (8, 12, 15) involve reading ability

Only one test assesses mechanical reasoning ability. Some tests appear to measure two or more abilities. Finally the remaining tests attempt to give some indication of memory, co-ordination and manipulative ability.

The sketch above indicates the similarities between the FACT battery and the DAT battery. The apparent advantage of the former battery is that it attempts to derive a greater number of ability measures appropriate to specific occupations.

Approximately seven hours and fifteen minutes are required to administer the complete battery, each test requiring about 20 minutes.[1] Raw scores have been converted into standardized (stanine) scores and norms are differentiated by sex and grade level (9-12).

Refinements have been made in terms of administration time and testing efficiency. The battery has been divided into two parts, each containing 14 tests; battery A purports to measure those abilities appropriate to occupations requiring college graduation and battery B,

---

1 While some tests are speeded, the greater segment of the FACT battery is comprised of power tests which allow over 90 per cent of the examinees to finish them. Tests 8 and 14 have no time limits.

those appropriate to occupations normally requiring high-school graduation. Batteries A and B each require about 5 hours to administer.

Many questions have been raised about FACT with respect to the quantity and quality of norms, reliability and validity data, and of inter-correlation values between tests. Data for these are insufficient, but the strong vocational emphasis in the design of the FACT battery makes it worthy of further study.

Holzinger-Crowder Uni-Factor Tests (1952-55)

This test is characterized by a special effort toward factorial purity; that is, factorial independence. There are four categories -- verbal, spatial, numerical, and reasoning, measured by nine subtests.

Separate sex norms are provided for each factor score and for each grade (7-12). Mitchell (1955) reports that "alternate-form reliability coefficients for the separate factor scores obtained separately for each grade (7, 9, 11), range from .76 to .95 with an average of .85". Validity data described in the same article appear acceptable for academic courses and certain vocational courses (Shop Mathematics, Science, Junior Business Training, and Mechanical Drawing). Data for several commercial courses (e. g. Typing, Shorthand, Accounting) were not as encouraging.

A valuable feature for the guidance counselor's use is the formula provided to compute a composite score which indicates a student's potential scholastic aptitude.

It appears that this battery must be supplemented by other tests to furnish differential ability measures.

Employee Aptitude Survey (1952-58)

The EAS battery, reviewed by Buros (1958), is composed of ten tests:

1. Verbal Comprehension
2. Numerical Ability
3. Visual Pursuit
4. Visual Speed and Accuracy
5. Space Visualization
6. Numerical Reasoning
7. Verbal Reasoning
8. Word Fluency
9. Manual Sp. and Acc.
10. Symbolic Reasoning

Although the battery was designed to predict occupational success, a large segment of the validity data is devoted to describing the correlations between test scores and in training courses. The remaining validity data based on occupations are again concurrent rather than predictive.

The Lockheed Aircraft Corporation has used the EAS extensively and results of their studies are published. The Manual for Interpreting the Employee Aptitude Survey supplies further normative data on occupational groups in the plant.

The EAS normally takes only an hour to administer, but the time element could be reduced if the Lockheed selection of tests were used (i. e., tests 1, 2, 4, 5, 6 and 7).

There is some suggestion that this battery tests only general learning ability. However, since Lockheed found this battery useful for job placement, it is possible that this battery might assist the counselor in directing students to appropriate programmes.

Criticism of this battery is aimed mainly at the method by which the authors derived their cut-off scores. Since no published explanation is given, one is led to assume that these values were determined on face validity or by "professional judgment".

IV SPECIALIZED TESTS

A multifactor test battery attempts to provide a set of aptitude tests which can be applied to a wide range of occupational choices. Its versatility, however, reduces the chances of accurate prediction. Given a specific vocational setting, the test designer has a greater knowledge of exactly what ability measures should be taken, and can take a greater sampling of these ability measures. In contrast, the multifactor test tends to be more abstract in content but less differentiating in isolating abilities.

Specialized tests may augment the differentiating capacity of the multifactor test. In the light of certain weaknesses generally found in multifactor batteries, an examination of some tests which offer ability measures apparently required of vocations in mechanical and clerical fields seems appropriate.

The Revised Minnesota Paper Form Board Test (1948)

In 20 minutes, this test attempts to measure mechanical aptitude and the related abilities generally required in mechanically oriented vocations. Aside from reading the simple instructions, no other reading or verbal reasoning ability is required. This is a particularly important feature where students with language handicaps are being tested. Although graduated in difficulty, the 64 items involve basically the same principle in visualizing and manipulating objects in space. Ability in perspective judgment and in discriminating size relations, and a capacity for attention to detail are required to succeed in this test.

The validation data in the manual suggest that the test is more successful at prediction in some occupational areas than in others. Concurrent validity correlations were relatively high for:

detail draftsmen	(.48)	
merchandise packer	(.49)	
inspector-packers	(.57)	
inspector	(.50)	engine and propeller industry
machine operators	(.38)	
foreman	(.47)	
job setters	(.59)	
tool room learners	(.44)	
senior dentists	(.61)	

Interform reliability of the test is fairly high (.85).

Scores for sex differences do not significantly differ. Only a slight distinction persists in that males consistently excel the females.

Two sets of norms are provided for each of two groups, educational and industrial. The authors point out, however, that the interpreter must select data which are most appropriate to his sample. The industrial group norms cover a wide age span.

Morgan (1944a, 1944b) reports findings very pertinent to aptitude testing in general as well as this test. Marks in elementary school were found useful in predicting junior high school shop grades. Scores on the Minnesota Paper Form Board Test (former edition) did not significantly differentiate the good Grade 10 from the good or poor Grade 12, nor did scores differentiate between the two Grade 12's. There was

also a consistent tendency for older pupils to get lower scores. The Revised Minnesota Paper Form Board Test was excluded from this testing programme on the grounds that the score range was too narrow, and did not provide adequate differentiation. A fundamental arithmetic test was adopted, and has been useful in selecting students for technical and industrial courses, and apprentices for industry.

This type of test might yield substantial correlations with vocational courses requiring spatial perception such as sewing, tailoring, printing, or other courses requiring visualizing of objects in space. The manual provides inadequate information on such topics nor is there information on the relation of this test to the DAT subtests. Correlations of the MPFB test with other tests of mechanical ability are generally low (-.09 to .60 with the median in the low .30's). This suggests that it may be testing a rather unique ability.

The Minnesota Clerical Test (1933-46)

The Minnesota Clerical Test, taking only 20 minutes to administer, has been one of the most popular tests of its kind. The testee is asked to compare digits, words, and names, and indicate whether or not they are identical.

This is a highly speeded test and low scores tend to reflect slow workers as well as careless workers. Response set, then, can alter a score.

Again, we can consider possible subject areas for which a clerical test might be appropriate, such as detail drafting, blueprint reading, printing, watch repair, or other fields that require attention to detail.

Turse Clerical Aptitudes Test (1955)

The Turse Clerical Aptitudes Test requires 40 minutes and attempts to measure not only clerical aptitude but also the capacity to learn processes involved in clerical occupations. Interpretation of this test is based on seven scores:

1. Verbal Skills
2. Number Skills
3. Written Directions

} aimed at measuring learning ability

4. Checking Speed
5. Classifying-Sorting
6. Alphabetizing

These six tests are combined to yield a "general clerical aptitude" score -- a composite of twice the learning score plus the clerical speed score.

Though separate sex norms are not provided, predictive validities and normative data reported are encouraging.

V APTITUDES VS. INTERESTS

The young student too often confuses high interest with high aptitude. Some guidance services have accepted an intelligence quotient and an interest profile as sufficient information to predict occupational success. Though interest and intelligence are important, adequate measures of special abilities would improve predictive success.

Experienced counselors are aware that certain interest-ability correlations recur often enough to discount pure coincidence. Factorial studies done by Smith (1958) point out that a relationship exists in some instances but not in a one-to-one ratio. Several independent measures, including the DAT, Kuder-Vocational and Kuder-Personal, did converge into various patterns. There has not as yet been either sufficient or conclusive evidence, however, to suggest that ability measures can be ignored or dispensed with in counseling.

Another problem in collecting data for guidance is the question of whether knowledge of his aptitude profile affects a student's interest profile; for example, in Kuder interest scores. Unfortunately, research evidence is contradictory. Two studies (Froehlich, 1954; Meek, 1954) claim that interest scores do change. Stewart's (1956) investigation yielded negative results. In any case, appropriate precautions can be taken to avoid contamination of Kuder scores.

Anderson (1953) contends that ability, interest, and personality measures must be examined to provide a balanced view of the student. There are many important qualities of the student not measured by interest or aptitude tests such as character, resourcefulness, initiative, or level of aspiration that may be related to success in school and work.

VI THE STUDENT AS A SOURCE OF ERROR IN PREDICTING ACHIEVEMENT

Most predictionsare conditional. One assumes that certain related factors will remain constant. For example, a student's aptitude score profile will not vary if he does his best on each successive retest, and provided that retests are far enough apart to eliminate practice effects.

Scholastic achievement is governed to some extent by the student's "level of awareness" or "reality orientation". He must be mature enough to resist short-term goals and outside influences (high parental aspirations, high preference ranks for some occupations, etc.).

Student motivational level can also be a guide in vocational prediction. Did the student do his best during testing? Will he strive to do his best in the advisable courses? That so many promising students with high chances of success do under-achieve, fail, or drop-out exemplifies the fact that vital areas of scholastic achievement have been ignored. Raph and Tannenbaum (1961) show clearly that even with the ideal aptitude test, many non-intellective factors can determine student success or failure. Decision-making in guidance and prediction must incorporate as many of these factors as possible if the level of confidence is to be increased.

These considerations indicate that the student can account for a great source of variability in testing and predicting success. Thus, while a test may be a good achievement predictor, it has its limitations to the extent that every aspect of the achievement process cannot be measured.

VII GENERAL CONCLUSION

Aptitude tests have proven to be useful instruments in predicting general scholastic aptitudes but they have not been as successful in predicting technical aptitudes. In spite of sophisticated test construction techniques, further research is required to identify and measure aptitudes more accurately. Mechanical and clerical tests are not, as yet, very successful instruments, having low correlations with success in relevant areas.

A major factor in a test's usefulness is the adequacy of its manual. Manuals frequently lack adequate norms, their source, and appropriate validity and reliability data. The DAT manual is an exception.

There is little probability that any of the present multifactor batteries will adequately differentiate aptitudes; the incorporation of other tests and information is needed for adequate guidance. Student motivation appears as one area requiring more attention and adequate measurement.

REFERENCES

Anastasi, Anne. Psychological testing New York: Brett-Macmillan Ltd., 1961.

Anderson, Rose, G. Do aptitudes support interests? Personnel guid. J., 1953, 32, 14-17.

Bennett, G. K., Seashore, H. G., and Wesman, A. G. Counseling from profiles: a casebook for the Differential Aptitude Tests. New York: Psychol. Corp., 1951.

Bennett, G. K., Seashore, H. G., and Wesman, A. G. "The differential aptitude tests: an overview". Comments by D. E. Super. Personnel guid. J., 1956, 35, 81-93.

Bennett, G. K., Seashore, H. G., and Wesman, A. G. Manual for the Differential Aptitude Tests (3rd ed.) New York: Psychol. Corp., 1959.

Berdie, R. F. The Differential Aptitude Tests as predictors in engineering training. J. educ. Psychol., 1951, 42, 114-123.

Buros, O. K., (ed.) The fifth mental measurements yearbook. New Jersey: Gryphon Press, 1959.

Flanagan, J. C. The Flanagan Aptitude Classification Tests. Personnel guid. J., 1957, 35, 495-507.

Froehlich, C. P. Does test taking change self ratings? Calif. J. educ. Res., 1954, 5, 166-169.

Meek, C. R. The effect of knowledge of aptitude upon interest scores. Unpublished Ph. D. dissertation. George Peabody College, 1954.

Mitchell, Blythe, C. "The relation of high school achievement to the abilities measured by the Holzinger-Crowder Unifactor Tests." Ed. psychol. Measmt., 1955, 15, 487-90.

Morgan, W. J. Some remarks and results of aptitude testing in technical and industrial schools. J. soc. Psychol., 1944 a, 20, 19-29.

Morgan, W. J. The scores on the Revised Minnesota Paper Form Board Test at different grade levels of a technical-industrial high school. J. genet. Psychol., 1944 b, 64, 159-162.

Raph, B. Jane and Tannenbaum, A. J. Underachievement: review of literature. Talented Youth Project, Horace Mann-Lincoln Institute of School Experimentation, Teachers' College, Columbia University, New York 27, New York, 1961. (Unpublished)

Smith, D. D. "Abilities and interests: I, a factorial study." Canad. J. Psychol., 1958, 12, 191-201.

Stewart, L. H. Does knowledge of performance on an aptitude test change scores on the Kuder? J. counsel. Psychol., 1957, 4, 161-64.

Vineyard, E. E. "A longitudinal study of the relationship of differential aptitude test scores with college success." Personnel guid. J., 1958, 36, 413-16.

Wolking, W. D. Predicting academic achievement with the differential aptitude and the primary mental abilities tests. J. appl. Psychol., 1955, 39, 115-118.